

PRANAV DHIRAN

Nagpur, Maharashtra | +91-8999629839 | dhiranpranav72@gmail.com | pranavdhiran.me | linkedin.com/in/prannav-dhiran | github.com/Pranav-d33

PROFILE

Engineer who ships end-to-end AI/ML systems: training and data pipelines, Python/Go backend services, and agentic infrastructure that runs in production. At AIISC (University of South Carolina) I own the pre-training pipeline for a 140M-parameter domain SLM — corpus processing at 6.5 GB scale, custom tokenizer, distributed training, and an automated evaluation harness. As an LFX '26 Mentee at Hyperledger Cello I build the Django/DRF backend, REST/SSE streaming APIs, and an LLM tool-calling agent; 21 PRs shipped. Also an open-source contributor to CNCF's Meshery/MeshKit (Go + React) and author of a published PyPI MCP server. Seeking an internship building production ML/LLM systems.

EXPERIENCE

Research Intern — AIISC, University of South Carolina Apr 2026 – Present

- Own the pre-training pipeline for a 5-person team building an India-focused agriculture SLM, from data collection and cleaning through pre-training and eval.
- Built a 140M-parameter Qwen3-style transformer training/eval pipeline (GQA, RoPE, RMSNorm, SwiGLU, factorized embeddings) on a 6.5 GB / 266K-document corpus, with training-health, domain-validation, and safety evaluation frameworks.
- Built the neurosymbolic layer: BPE tokenizer with AGROVOC entity injection, Z3-based causal KG verification, Triple Transformer Encoder.
- Built India-Agri-KG, a deterministically-verified agricultural knowledge base (784 districts, 26 domains, 14,478+ entities, 38,590+ relations) with a 5-stage non-LLM verification layer enforcing source traceability on every triple (unpublished).
- Delivered two technical lectures for the IAIRO-affiliated SLM Bootcamp 2026: [Vanilla GPT-2 Architecture](#) and [Scaling Laws & Cost Accounting](#).

LFX Mentee — Hyperledger Cello · Linux Foundation Jun 2026 – Present

- Building an AI operations copilot for Hyperledger Fabric, connecting natural-language queries to Cello REST APIs through an LLM tool-calling workflow.
- Implemented the Django/DRF backend and SSE streaming pipeline powering incremental AI responses in the React dashboard.
- Built a Fabric node-logs API on the agent layer with bounded log retrieval and Docker error handling for AI-assisted node debugging.
- Developed a multi-party channel invitation workflow end to end: database models, REST APIs, Fabric configuration/signing logic, and dashboard UI.
- Contributed 21 PRs (11 merged) spanning feature development, debugging, CI/test fixes, dependency updates, and API/data-model fixes.

Open Source Contributor — Meshery, CNCF Sandbox Project Mar 2026 – Jul 2026

- Contributed 5+ merged PRs across Meshery's Go backend and React frontend: service-mesh features, UI components, and API integrations.
- Active in issue triage, code reviews, and contributor discussions under CNCF's open-source workflow.

PROJECTS

GNU Radio MCP Server — LLM-to-SDR Bridge Python · FastMCP · ZMQ · XML-RPC · GNU Radio [pypi](#) | [github](#)

- Built and published a provider-agnostic MCP server to PyPI bridging LLMs to live GNU Radio SDR flowgraphs over stdio transport.
- Implemented 13 tools for flowgraph lifecycle, frequency/gain control, and signal monitoring, with Pydantic v2 validation, ZMQ IQ streaming (Welch PSD), and an XML-RPC control plane for natural-language control of HackRF One and RTL-SDR hardware via Claude Desktop.

Medaura — Agentic Pharmacy System FastAPI · LangChain · ChromaDB · Langfuse · React [live](#) | [github](#)

- Built a 5-agent autonomous pharmacy system (Ordering, Safety, Forecast, Procurement, UI) with semantic RAG via ChromaDB and Langfuse observability — live on the web, zero human intervention, 4 languages.

Small Language Model from Scratch PyTorch · Hugging Face · AMP · NLP [github](#)

- Pre-trained a GPT-style autoregressive transformer (6 layers, 6 heads, 384-dim) on TinyStories with a custom BPE tokenizer, mmap data pipelines, AMP mixed precision, gradient accumulation, and cosine LR scheduling; ran depth/embedding ablations to characterize scaling behavior.

RF Watch — Open-Source RF Spectrum Monitor Python · GNU Radio · HackRF One · DSP [github](#)

- Built a real-time RF spectrum analyzer with FFT-based feature extraction and a lightweight ML classifier for passive transmitter detection, built out of anti-drone work for ITBP (SIH 2025).

TECHNICAL SKILLS

Languages & Frameworks: Python, Go, PyTorch, TensorFlow, Hugging Face Transformers, TRL

APIs & Protocols: REST, gRPC, GraphQL, OpenAI/Gemini/Ollama APIs, FastMCP, XML-RPC, ZMQ

LLM Engineering: Instruction Fine-Tuning, LoRA, PEFT, GRPO, RLHF, INT4/INT8 Quantization, Unsloth, LangChain, LangGraph, ChromaDB, FAISS, RAG

Agentic & Infra: Multi-Agent Systems, Tool-Use, Function Calling, MCP Servers, LangSmith, Langfuse, Docker (basic), Kubernetes (basic)

EDUCATION

B.Tech — Electronics & Telecommunication Engineering (Minor in IT) 2023 – 2027

SGGS Institute of Engineering & Technology, Nanded

Coursework: Discrete Mathematics, Probability & Statistics, Engineering Mathematics (Calculus & Linear Algebra), Signals & Systems, Digital Signal Processing, Digital Electronics, Communication Systems, Data Structures & Algorithms, Database Management Systems

AWARDS & RECOGNITION

- Qualified — ETHGlobal 2026
- International Finalist (Top 6) — UWA Hack 2026
- National Finalist — SIH 2024 & 2025
- Regional Qualifier — Nxt Wave x OpenAI Buildathon

CERTIFICATIONS

- Fine-tuning & RL for LLMs: Intro to Post-training — DeepLearning.AI
- AI Agents in LangGraph — DeepLearning.AI
- Quantization Fundamentals — Hugging Face
- MCP: Build Rich-Context AI Apps — Anthropic